

# PBmice: an integrated database system of *piggyBac* (PB) insertional mutations and their characterizations in mice

Ling V. Sun<sup>1,2,\*</sup>, Ke Jin<sup>2</sup>, Yiming Liu<sup>3</sup>, Wenwei Yang<sup>1,2</sup>, Xing Xie<sup>3</sup>, Lin Ye<sup>2</sup>, Li Wang<sup>2,4</sup>, Lin Zhu<sup>3</sup>, Sheng Ding<sup>1,2</sup>, Yi Su<sup>3</sup>, Jie Zhou<sup>2</sup>, Min Han<sup>1,2,5</sup>, Yuan Zhuang<sup>1,2,6</sup>, Tian Xu<sup>1,2,7</sup>, Xiaohui Wu<sup>1,2</sup>, Ning Gu<sup>3,\*</sup> and Yang Zhong<sup>2,4,\*</sup>

<sup>1</sup>Institute of Developmental Biology and Molecular Medicine, <sup>2</sup>School of Life Sciences, <sup>3</sup>School of Information Science and Engineering, Fudan University, Shanghai 200433, <sup>4</sup>Shanghai Center for Bioinformation Technology, Shanghai 200235, China, <sup>5</sup>Howard Hughes Medical Institute, Department of MCDB, University of Colorado at Boulder, Boulder, CO 80309, <sup>6</sup>Department of Immunology, Duke University Medical Center, Durham, NC 27710 and <sup>7</sup>Howard Hughes Medical Institute, Department of Genetics, Yale University School of Medicine, Boyer Center for Molecular Medicine, New Haven, CT 06536, USA

Received August 13, 2007; Revised September 13, 2007; Accepted September 17, 2007

## ABSTRACT

DNA transposon *piggyBac* (PB) is a newly established mutagen for large-scale mutagenesis in mice. We have designed and implemented an integrated database system called PBmice (PB Mutagenesis Information Center) for storing, retrieving and displaying the information derived from PB insertions (INSERTs) in the mouse genome. This system is centered on INSERTs with information including their genomic locations and flanking genomic sequences, the expression levels of the hit genes, and the expression patterns of the trapped genes if a trapping vector was used. It also archives mouse phenotyping data linked to INSERTs, and allows users to conduct quick and advanced searches for genotypic and phenotypic information relevant to a particular or a set of INSERT(s). Sequence-based information can be cross-referenced with other genomic databases such as Ensembl, BLAST and GBrowse tools used in PBmice offer enhanced search and display for additional information relevant to INSERTs. The total number and genomic distribution of PB INSERTs, as well as the availability of each PB insertional LINE can also be viewed with user-friendly interfaces. PBmice is freely available at <http://www.idmshanghai.cn/PBmice> or <http://www.scbt.org/PBmice/>.

## INTRODUCTION

Large-scale mutagenesis is critical to functional characterization of the mammalian genomes. In mice, gene targeting, chemical mutagenesis and insertional mutagenesis are available to achieve this goal. The International Knockout Mouse Consortium (IKMC) has recently been formed to systematically produce null alleles for individual genes in mice (1–3). This approach focuses on annotated protein-coding regions but provides limited insight into the rest of the genome. In contrast, ENU (*N*-ethyl-*N*-nitrosourea) based chemical mutagenesis has the capacity to introduce mutations throughout the entire genome (4–11). Several large-scale ENU mutagenesis studies have been conducted in the recent years and produced a large collection of mutant mice with interesting phenotypes. However, the majority of these mutant strains remain to be mapped to the single gene level. Insertional mutagenesis has been shown to be one of the most efficient means to generate a large number of identifiable mutants throughout the entire genome of several model organisms (12–17). With an appropriate insertion vector, insertional mutagenesis can produce a large number of mutations at low cost with a high speed. Each insertion site is tagged with the vector sequence and thus can be easily mapped.

Retroviruses have been used in mouse genetics for more than two decades (18,19). They led to discovery of a number of proto-oncogene and tumor suppressor loci (19). However, successful application of retroviruses in

\*To whom correspondence should be addressed. Tel: +86 21 55664436; Fax: +86 21 65642468; Email: yangzhong@fudan.edu.cn  
Correspondence may be addressed to Ling V. Sun – Tel: +86 21 65642244; Fax: +86 21 65643770; Email: lingsun@fudan.edu.cn and Ning Gu – Tel: +86 21 55664462; Fax: +86 21 55664456; Email: ninggu@fudan.edu.cn

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© 2007 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

large-scale insertional mutagenesis remains limited to gene trapping in ES cells (20,21). In recent years, the progress of introducing transposon tools into mammalian genetic studies has offered new opportunities for mutagenesis in mice (22–30). Several transposons, such as *Sleeping Beauty* (SB), *Minos*, *Tol2*, *Mos1* and *piggyBac* (PB), were shown to be active in mammalian cells and/or in mice. Among them, the PB transposon appears to be a promising option. Appreciation of PB as an important genetic manipulation tool in various organisms including mice has been growing rapidly in recent years, as accumulating evidence has indicated that PB elements are capable of efficient transposition in animals from diverse genera of insects, flatworms, mice, as well as mammalian cell lines (28,29,31–38). PB elements, originally found in the cabbage looper moth *Trichoplusia ni*, are DNA transposons carrying a unique functional transposase (PBase) of 594 amino acids (39–41). It has been reported to be more active than *Tol2*, *Mos1* and SB in mammalian cells (30), and has been shown to be effective mutagen for insertional mutagenesis in mice.

Currently, a genome-wide PB mutagenesis project is underway in the Institute of Developmental Biology and Molecular Medicine (IDM) at Fudan University in Shanghai, China. To collect and disseminate the large amount of genetic and phenotypic information from this project, we have designed and implemented an integrated database system called PBmice (PB Mutagenesis Information Center), which provides a user interface to query and display data obtained from PB insertions (INSERTs) in the mouse genome and their characterizations.

## SYSTEM DESIGN AND IMPLEMENTATION

### Data sources

Experimental data in PBmice are derived from the on-going project of PB insertional mutagenesis in mice at IDM. Each PB transposition event produces a PB insertion, which is termed as an INSERT in PBmice and carried by a mouse line (LINE). The flanking genomic sequence of an INSERT is characterized by inverse PCR and determined for its location in the mouse genome by WU-BLAST 2.0 search against the Ensembl mouse genome sequences (Ensembl release 45, based on GenBank m36 mouse assembly, [http://www.ensembl.org/Mus\\_musculus/index.html](http://www.ensembl.org/Mus_musculus/index.html)) (42). Expression level of the gene hit by an INSERT is determined by real-time RT-PCR. Other information about the INSERT, such as the mouse strain and LINE carrying the INSERT and the PB construct used to generate the INSERT, are also collected in the database. When a trapping vector is used to generate the INSERT, the expression pattern of the reporter gene of the vector, an indication of the expression pattern of the endogenous gene affected in the INSERT, is also examined and collected in the database. Associated information of a LINE carrying one or more INSERTs includes the existence of live animals and storage status, as well as its phenotyping data. When an INSERT hits a functional genomic unit, such as an annotated protein-coding gene, a gene for a microRNA, or a regulatory

element, this interruption of genomic function may be shown as phenotypes in the hosting LINE. Such characterizations are also recorded and documented in PBmice. In addition, other information, such as gene annotation data, is downloaded from Ensembl and GenBank.

### Data integration

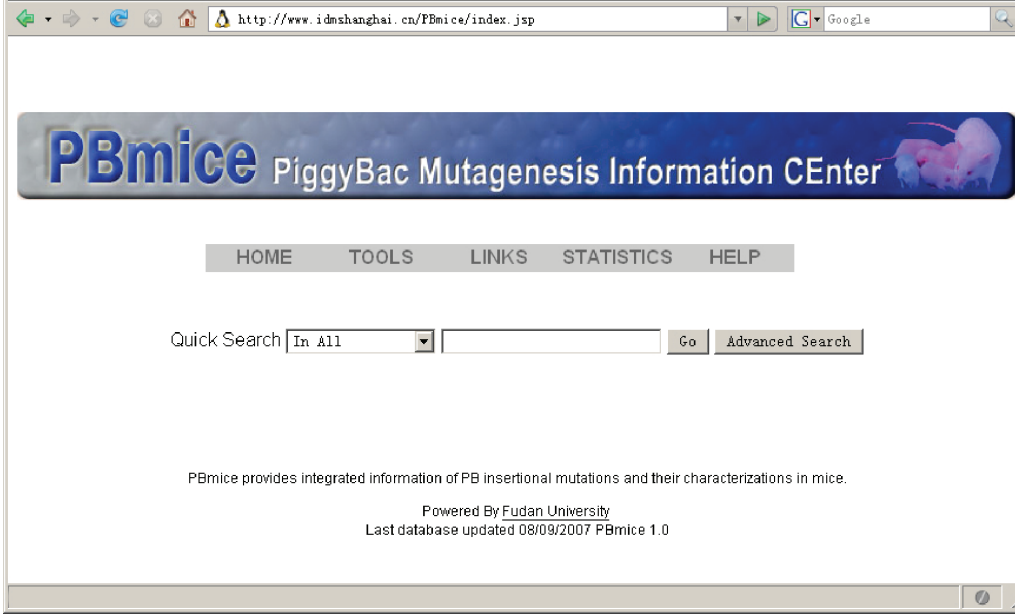
A Java tool has been developed for PBmice to integrate gene data from Ensembl and GenBank are incorporated together. MGI-ID of a gene is used to integrate the information of the same gene downloaded from Ensembl and GenBank in PBmice. Gene information contains gene names with synonyms, descriptions, genomic locations, transcript information and IDs linked to existing databases, such as Ensembl, GenBank and Mouse Genome Informatics (MGI) (<http://www.informatics.jax.org/>), for links to its detailed information in those databases. Additionally, users can obtain protein information from ExPASy (<http://us.expasy.org/>) and gene function information from the Gene Ontology Database (<http://www.geneontology.org/>). The integrated items are converted into data files in GFF format (<http://www.sanger.ac.uk/Software/formats/GFF/>) for retrieving and displaying with GBrowse (<http://www.gmod.org/wiki/index.php/GBrowse>), allowing for a more straightforward and user-friendly view.

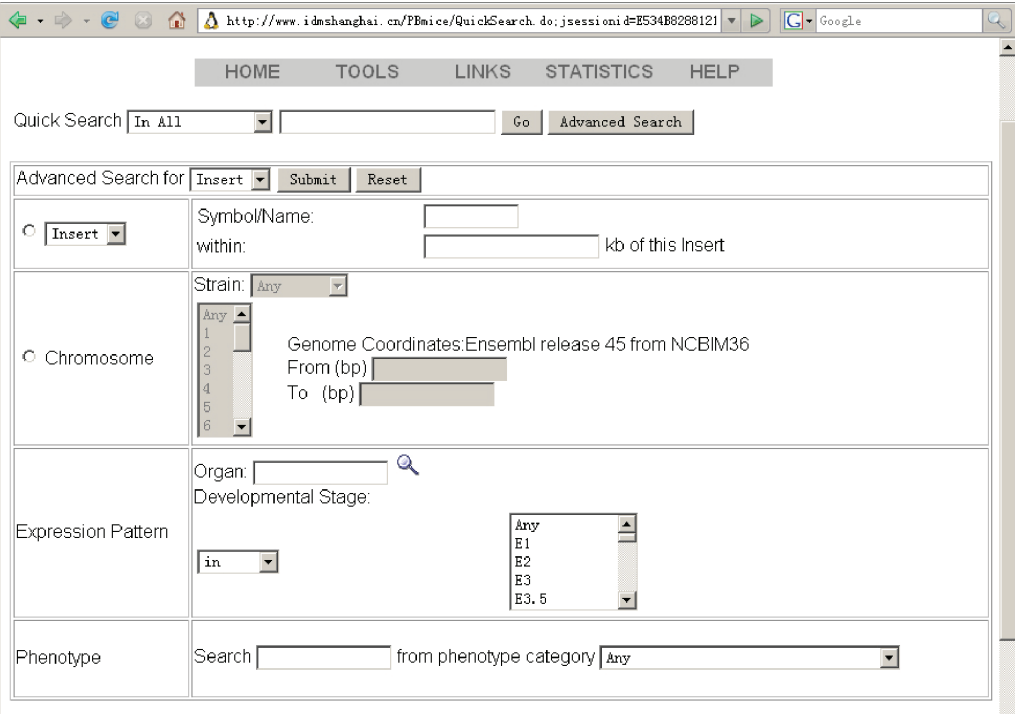
### User interface

INSERTs and INSERT-related data form the central core of PBmice. This system provides detailed information about an INSERT and associated information of the LINE carrying the INSERT and the gene hit by the INSERT.

To search for an INSERT of interest, either the quick search or the advanced search method can be used (Figure 1). The quick search allows a user to conduct a simple search just with the name of an INSERT, a LINE, a strain or a phenotype. A default entry results in a complete list of all INSERTs. The partial-word-match method is embedded in the query engine as well. Advanced search is provided for users with more defined interests to narrow down the search area, such as (i) to search for the existence of INSERTs in terms of their relative positions to a known INSERT or Gene; (ii) to search for INSERTs in a certain genomic area on a chromosome of interest; (iii) to search for INSERTs associated with the expression pattern characterized for a certain organ in an animal of certain developmental stage or (iv) to search for INSERTs associated with a certain phenotype. The first and second search criteria can be independently combined with the rest of search criteria but not with each other during an advanced search.

The *INSERT Detail* interface presents the detailed information of an INSERT after a successful search, including chromosomal location and flanking genomic sequence of the INSERT, the expression level of the gene that is hit by the INSERT, and the PB construct used to generate the INSERT (Figure 2). Of all the information displayed in the interface, the *Chromosome Location Prox* item refers to the genomic position immediately preceding

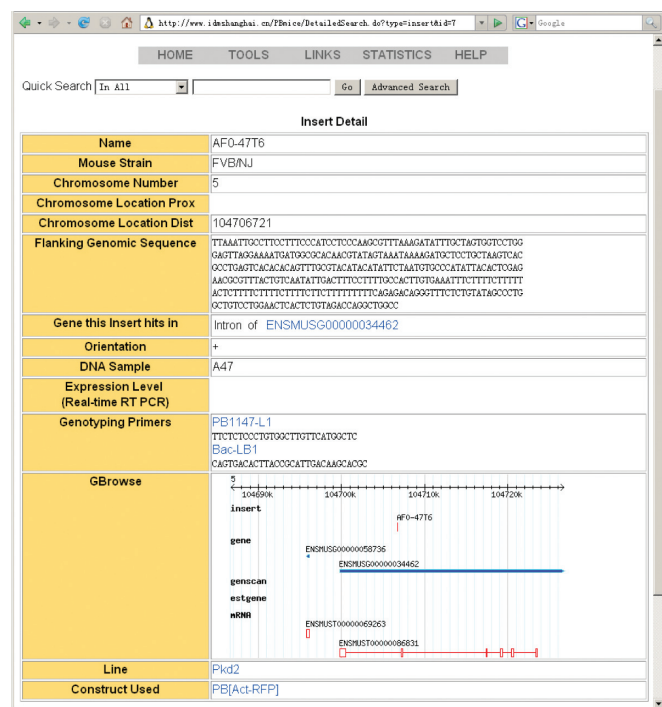
(a) 

(b) 

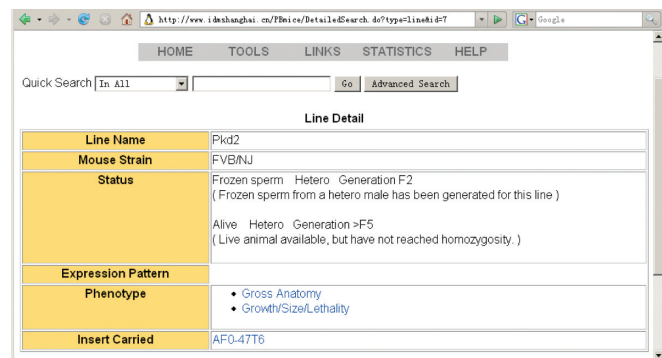
**Figure 1.** Search interfaces. (a) The *Quick Search* interface. This is also the homepage of PBmice. (b) The *Advanced Search* interface.

the INSERT, while the *Chromosome Location Dist* item refers to the position immediately following the INSERT. The *Gene hit* item points to the information of the gene with its Ensembl ID, which is clickable for users to obtain detailed information of this gene. The *Orientation* item of this INSERT is defined as following: the orientation of PB is defined as from PBL to PBR (29), while the chromosome orientation is defined by Ensembl. If the orientation of PB is the same as the orientation of the inserted chromosome, the orientation of the INSERT is denoted as '+', otherwise, the orientation of this INSERT will be '-'.

The *DNA Sample* item shows the name of the DNA sample used for mapping the INSERT. The *Expression Level* item indicates the expression level of the gene hit by the INSERT in the mutant. The *GBrowse* item offers an annotated view of a 40 kb genomic fragment with the position of the INSERT in the middle and is linked to the GBrowse tool embedded in PBmice for a user-adjustable view of the annotation of the genomic fragment around the INSERT. The *LINE* item contains the name of the mouse LINE that harbors this INSERT. The rest of the information categories titles are self-explanatory.

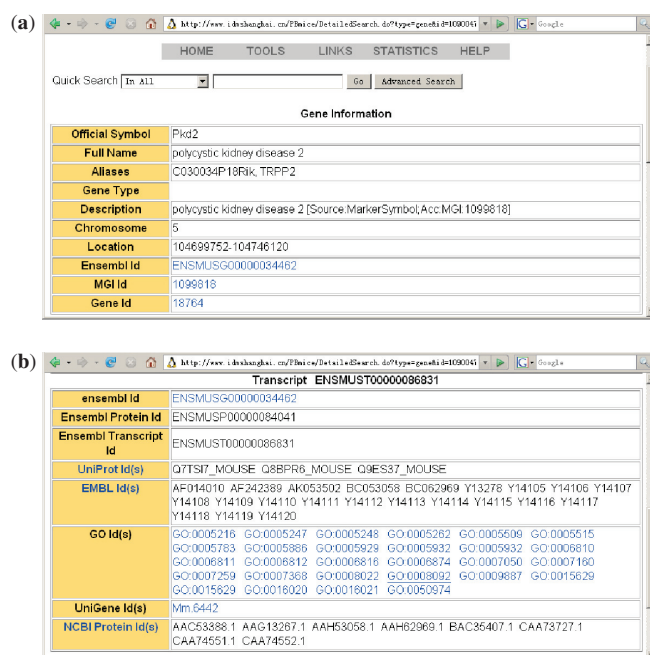


**Figure 2.** An example of the *INSERT Detail* interface. Information is based on INSERT AF0-47T6 in PBmice.



**Figure 3.** An example of the *LINE* interface. Information is based on LINE Pkd2.

In the PBmice database, the information of a LINE is associated with the INSERT(s) it carries. The *LINE Detail* interface reveals the properties of a LINE carrying one or more INSERTs, including the status, expression pattern and phenotyping data of the LINE (Figure 3). In this interface, the *Status* item indicates the storage status of the LINE (live animal, frozen sperm and/or frozen embryo) and related information for strain distribution. When a LINE carries trapping vector-related INSERT(s), the *Expression pattern* item will indicate any information that was characterized to some extent for the expression of the reporter protein. The *Phenotype* item provides clickable categories of phenotypes characterized for this LINE. The detailed phenotypic data contains not only the description of the phenotype and assays used to depict the phenotype, but also the sample information, such as



**Figure 4.** Examples of the *Gene Information* interface and *Transcript* interface. Data displayed is based on the Pkd2 gene.

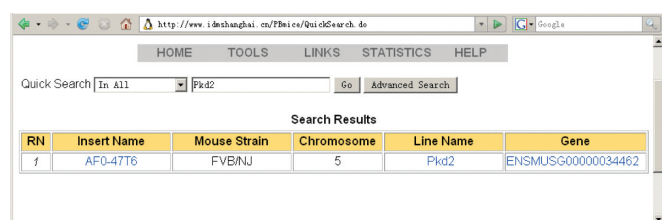
the heterozygosity or homozygosity, the developmental stage, the sex of the animals or the organ examined. Users can also retrieve related INSERTs/LINEs based on similarity of phenotypes through advanced search (Figure 1b).

PBmice also provides information of a gene when hit by one or more INSERT(s). Both the *Gene Information* interface and the *Transcript* interface exhibit the identity and associated information of a gene (Figure 4a and b). The *Gene Information* interface presents the name and synonyms, the description and the genomic location of a gene and the links to outside databases, such as Ensembl, MGI and GenBank (Figure 4a). The *Transcript* interface displays links to existing databases for users to retrieve detailed information about each related transcript, protein and gene function information of the gene as shown in Figure 4a (Figure 4b). UniProt-ID(s), EMBL-ID(s) and GenBank-Protein-ID(s) are linked to information about the protein expressed by this gene stored in ExPASy containing UniProt Knowledge base (Swiss-Prot and TrEMBL) through the Multi-Protein Survey System (MPSS) developed by Bioinformatics Center, Shanghai Institutes for Biological Sciences (<http://www.biosino.org/MPSS/index.jsp>). GO-ID(s) are linked to the associated gene function information from the Gene Ontology database. UNIGENE-ID(s) are linked to the associated transcripts information from GenBank.

### Example query

Taking the INSERT AF0-47T6 that hits into the gene *Pkd2* as an example to navigate PBmice. *Pkd2* was cloned in 1996 and proved to contribute to autosomal dominant polycystic kidney disease (ADPKD) when mutated (43). Patients suffering from ADPKD grow cysts in their





The screenshot shows a web browser window with the URL <http://www.idmshanghai.cn/PBmice/QuickSearch.do>. The page has a navigation bar with links: HOME, TOOLS, LINKS, STATISTICS, and HELP. Below the navigation bar is a search box with the text 'Quick Search' and a dropdown menu set to 'In All'. The search term 'Pkd2' is entered in the search box, and a 'Go' button is visible. Below the search box is a 'Search Results' table with the following data:

RN	Insert Name	Mouse Strain	Chromosome	Line Name	Gene
1	AF0-47T6	FVB/NJ	5	Pkd2	ENSMUSG00000034462

**Figure 5.** An example of the *Search Result* interface, using search term *Pkd2*.

kidneys eventually leading to kidney failure. ADPKD is the most common inherited disease affecting about half a million Americans and more than 1.5 million Chinese (44,45). To search for any INSERT existing that hits *Pkd2*, either quick search or advanced search can be used (Figure 1). Input *Pkd2* in the search box in the *Quick Search* interface (Figure 1a) will lead to a *Search Result* interface displaying the following information (Figure 5): (i) INSERT Name AF0-47T6, which is clickable to display the *INSERT Detail* interface for AF0-47T6 (Figure 2); (ii) Mouse strain FVB/NJ carrying AF0-47T6; (iii) Chromosome 5 on which AF0-47T6 is located; (iv) LINE Name *Pkd2* carrying AF0-47T6, which is also a link to exhibit the *LINE Detail* interface for *Pkd2* (Figure 3) and (v) Gene hit by AF0-47T6 in its Ensembl ID ENSMUSG00000034462, which leads to the *Gene Information and Transcript* interface for the *Pkd2* gene (Figure 4a and b). In the *LINE Detail* interface for *Pkd2*, two types of phenotypes have been characterized for this LINE, including detailed information that LINE *Pkd2* has phenotypes similar to those of the previously described *Pkd2* knockout mice (29,46). Advanced Search can also be used to retrieve any INSERT relevant to *Pkd2* (Figure 1b). Currently, three options are available in the *Advanced Search* interface to reach this goal: (i) search for any INSERT within 0 kb of the Gene *Pkd2*; (ii) search for any INSERT on Chromosome 5 within 104 699 752–104 746 120 bp and (iii) search for any INSERT associated with any phenotype falling into either category Gross Anatomy or category Growth/Size/Lethality. Submission of the query using option 1 will lead to an intermediate interface to let users choose the right gene of interest to continue, because more than one gene name contains the four characters, *pkd2*. Submission of the query using options 2 or 3 directly leads to a *Search Result* interface similar as shown in Figure 5.

BLAST and GBrowse tools are provided for enhanced search and user-friendly access to additional information relevant to each INSERT. The BLAST tool allows users access to any INSERT that hits into a mouse genomic DNA fragment of interest with the sequence information of this DNA fragment. The GBrowse tool offers users not only a simple and direct view of the genomic location of any INSERT in PBmice and the annotation of the mouse genome with user-defined criteria, but also links to the *INSERT Detail* interface in PBmice and the *Gene Report* interface in Ensembl. The *Statistics* interface presents the total number and distribution of INSERTs on the mouse chromosomes. A pie chart categorizes the genomic locations of all INSERTs into introns, exons, intergenic

regions or ND (undetermined). Introduction and step-by-step search examples are provided in the *Help* interface of PBmice.

## AVAILABILITY AND FUTURE DIRECTIONS

PBmice is freely available at <http://www.idmshanghai.cn/PBmice>, or <http://www.scbt.org/PBmice/>. It is maintained by the IDM at Fudan University. All data sources used by PBmice are reviewed and updated quarterly. During the next 3 years, development of this system will include: (i) new INSERTs will be released from the on-going genome-wide PB mutagenesis project that aims to create 10 000 insertions in mice; (ii) corresponding changes will be made following any substantial change in the mouse genome sequence information from new releases of linked databases such as Ensembl. In particular, the position of the existing INSERTs will be remapped with a new round of Blast search to provide updated information for users; (iii) a dynamic data management system will be incorporated into PBmice so that collaborators have the opportunities to access to primary data prior to its release and (iv) a policy and practical mechanism will be developed to allow dissemination of mutant strains and relevant reagents for non-profit academic research.

## ACKNOWLEDGEMENTS

This research was supported by grants from Chinese Key Projects for Basic Research (973) (Grant No. 2006CB806700, 2005CB321905 and 2002CB512801), Hi-tech Research and Development Project (863) (Grant No. 2007AA022012), National Natural Science Foundation of China (Grant No. 60473124), Shanghai Pujiang Program (Grant No. 05PJ14024) and Program for New Century Excellent Talents in University. We thank Rick Yan from Plano, Texas (TASM) for reading the early version of the manuscript. Funding to pay the Open Access publication charges for this article was provided by Hi-tech Research and Development Project (863) (Grant No. 2007AA022012).

*Conflict of interest statement.* None declared.

## REFERENCES

1. Austin, C.P., Battey, J.F., Bradley, A., Bucan, M., Capecchi, M., Collins, F.S., Dove, W.F., Duyk, G., Dymecki, S. *et al.* (2004) The knockout mouse project. *Nat. Genet.*, **36**, 921–924.
2. Auwerx, J., Avner, P., Baldock, R., Ballabio, A., Balling, R., Barbacid, M., Berns, A., Bradley, A., Brown, S. *et al.* (2004) The European dimension for the mouse genome mutagenesis program. *Nat. Genet.*, **36**, 925–927.
3. Collins, F.S., Rossant, J. and Wurst, W. (2007) A mouse for all reasons. *Cell*, **128**, 9–13.
4. Clark, A.T., Goldowitz, D., Takahashi, J.S., Vitaterna, M.H., Siepka, S.M., Peters, L.L., Frankel, W.N., Carlson, G.A., Rossant, J. *et al.* (2004) Implementing large-scale ENU mutagenesis screens in North America. *Genetica*, **122**, 51–64.
5. Fahrer, A.M., Bazan, J.F., Papathanasiou, P., Nelms, K.A. and Goodnow, C.C. (2001) A genomic view of immunology. *Nature*, **409**, 836–838.
6. Hrabe de Angelis, M.H., Flaswinkel, H., Fuchs, H., Rathkolb, B., Soewarto, D., Marschall, S., Heffner, S., Pargent, W., Wuensch, K.

- et al.* (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat. Genet.*, **25**, 444–447.
7. Nolan, P.M., Peters, J., Strivens, M., Rogers, D., Hagan, J., Spurr, N., Gray, I.C., Vitor, L., Brooker, D. *et al.* (2000) A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat. Genet.*, **25**, 440–443.
  8. Nolan, P.M., Peters, J., Vitor, L., Strivens, M., Washbourne, R., Hough, T., Wells, C., Glenister, P., Thornton, C. *et al.* (2000) Implementation of a large-scale ENU mutagenesis program: towards increasing the mouse mutant resource. *Mamm. Genome*, **11**, 500–506.
  9. Pargent, W., Heffner, S., Schable, K.F., Soewarto, D., Fuchs, H. and Hrabe de Angelis, M. (2000) MouseNet database: digital management of a large-scale mutagenesis project. *Mamm. Genome*, **11**, 590–593.
  10. Rathkolb, B., Decker, T., Fuchs, E., Soewarto, D., Fella, C., Heffner, S., Pargent, W., Wanke, R., Balling, R. *et al.* (2000) The clinical-chemical screen in the Munich ENU Mouse Mutagenesis Project: screening for clinically relevant phenotypes. *Mamm. Genome*, **11**, 543–546.
  11. Soewarto, D., Fella, C., Teubner, A., Rathkolb, B., Pargent, W., Heffner, S., Marschall, S., Wolf, E., Balling, R. *et al.* (2000) The large-scale Munich ENU-mouse-mutagenesis screen. *Mamm. Genome*, **11**, 507–510.
  12. Cooley, L., Kelley, R. and Spradling, A. (1988) Insertional mutagenesis of the *Drosophila* genome with single P elements. *Science*, **239**, 1121–1128.
  13. Zhang, P. and Spradling, A.C. (1994) Insertional mutagenesis of *Drosophila* heterochromatin with single P elements. *Proc. Natl Acad. Sci. USA*, **91**, 3539–3543.
  14. Amsterdam, A., Burgess, S., Golling, G., Chen, W., Sun, Z., Townsend, K., Farrington, S., Haldi, M. and Hopkins, N. (1999) A large-scale insertional mutagenesis screen in zebrafish. *Genes Dev.*, **13**, 2713–2724.
  15. Amsterdam, A., Yoon, C., Allende, M., Becker, T., Kawakami, K., Burgess, S., Gaiano, N. and Hopkins, N. (1997) Retrovirus-mediated insertional mutagenesis in zebrafish and identification of a molecular marker for embryonic germ cells. *Cold Spring Harb. Symp. Quant. Biol.*, **62**, 437–450.
  16. Koiwa, H., Bressan, R.A. and Hasegawa, P.M. (2006) Identification of plant stress-responsive determinants in *Arabidopsis* by large-scale forward genetic screens. *J. Exp. Bot.*, **57**, 1119–1128.
  17. Kuromori, T., Wada, T., Kamiya, A., Yaguchi, M., Yokouchi, T., Imura, Y., Takabe, H., Sakurai, T., Akiyama, K. *et al.* (2006) A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of *Arabidopsis*. *Plant J.*, **47**, 640–651.
  18. Varmus, H.E. (1983) Using retroviruses as insertional mutagens to identify cellular oncogenes. *Prog. Clin. Biol. Res.*, **119**, 23–35.
  19. Uren, A.G., Kool, J., Berns, A. and van Lohuizen, M. (2005) Retroviral insertional mutagenesis: past, present and future. *Oncogene*, **24**, 7656–7672.
  20. Stanford, W.L., Cohn, J.B. and Cordes, S.P. (2001) Gene-trap mutagenesis: past, present and beyond. *Nat. Rev. Genet.*, **2**, 756–768.
  21. Schnutgen, F., De-Zolt, S., Van Sloun, P., Hollatz, M., Floss, T., Hansen, J., Altschmied, J., Seisenberger, C., Ghyselinck, N.B. *et al.* (2005) Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proc. Natl Acad. Sci. USA*, **102**, 7221–7226.
  22. Ivics, Z., Kaufman, C.D., Zayed, H., Miskey, C., Walisko, O. and Izsvak, Z. (2004) The sleeping beauty transposable element: evolution, regulation and genetic applications. *Curr. Issues Mol. Biol.*, **6**, 43–55.
  23. Luo, G., Ivics, Z., Izsvak, Z. and Bradley, A. (1998) Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA*, **95**, 10769–10773.
  24. Dupuy, A.J., Fritz, S. and Largaespada, D.A. (2001) Transposition and gene disruption in the male germline of the mouse. *Genesis*, **30**, 82–88.
  25. Fischer, S.E., Wienholds, E. and Plasterk, R.H. (2001) Regulated transposition of a fish transposon in the mouse germ line. *Proc. Natl Acad. Sci. USA*, **98**, 6759–6764.
  26. Horie, K., Yusa, K., Yae, K., Odajima, J., Fischer, S.E., Keng, V.W., Hayakawa, T., Mizuno, S., Kondoh, G. *et al.* (2003) Characterization of sleeping beauty transposition and its application to genetic screening in mice. *Mol. Cell. Biol.*, **23**, 9189–9207.
  27. Zagoraiou, L., Drabek, D., Alexaki, S., Guy, J.A., Klinakis, A.G., Langeveld, A., Skavdis, G., Mamalaki, C., Grosveld, F. *et al.* (2001) In vivo transposition of Minos, a *Drosophila* mobile element, in mammalian tissues. *Proc. Natl Acad. Sci. USA*, **98**, 11474–11478.
  28. Wu, S., Ying, G., Wu, Q. and Capecchi, M.R. (2007) Toward simpler and faster genome-wide mutagenesis in mice. *Nat. Genet.*, **39**, 922–930.
  29. Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y. and Xu, T. (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell*, **122**, 473–483.
  30. Wu, S.C., Meir, Y.J., Coates, C.J., Handler, A.M., Pelczar, P., Moisyadi, S. and Kaminski, J.M. (2006) piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc. Natl Acad. Sci. USA*, **103**, 15008–15013.
  31. Morales, M.E., Mann, V.H., Kines, K.J., Gobert, G.N., Fraser, M.J. Jr, Kalinna, B.H., Correnti, J.M., Pearce, E.J. and Brindley, P.J. (2007) piggyBac transposon mediated transgenesis of the human blood fluke, *schistosoma mansoni*. *FASEB J.* doi: 10.1096/fj.07-8726com [Epub ahead of print].
  32. Sarkar, A., Sim, C., Hong, Y.S., Hogan, J.R., Fraser, M.J., Robertson, H.M. and Collins, F.H. (2003) Molecular evolutionary analysis of the widespread piggyBac transposon family and related “domesticated” sequences. *Mol. Genet. Genomics*, **270**, 173–180.
  33. Gonzalez-Estevez, C., Momose, T., Gehring, W.J. and Salo, E. (2003) Transgenic planarian lines obtained by electroporation using transposon-derived vectors and an eye-specific GFP marker. *Proc. Natl Acad. Sci. USA*, **100**, 14046–14051.
  34. Lorenzen, M.D., Berghammer, A.J., Brown, S.J., Denell, R.E., Klingler, M. and Beeman, R.W. (2003) piggyBac-mediated germline transformation in the beetle *Tribolium castaneum*. *Insect Mol. Biol.*, **12**, 433–440.
  35. Thibault, S.T., Singer, M.A., Miyazaki, W.Y., Milash, B., Dompe, N.A., Singh, C.M., Buchholz, R., Demsky, M., Fawcett, R. *et al.* (2004) A complementary transposon tool kit for *Drosophila melanogaster* using P and piggyBac. *Nat. Genet.*, **36**, 283–287.
  36. Sumitani, M., Yamamoto, D.S., Oishi, K., Lee, J.M. and Hatakeyama, M. (2003) Germline transformation of the sawfly, *Athalia rosae* (Hymenoptera: Symphyta), mediated by a piggyBac-derived vector. *Insect Biochem. Mol. Biol.*, **33**, 449–458.
  37. Handler, A.M. (2002) Use of the piggyBac transposon for germ-line transformation of insects. *Insect Biochem. Mol. Biol.*, **32**, 1211–1220.
  38. Balu, B., Shoue, D.A., Fraser, M.J. Jr and Adams, J.H. (2005) High-efficiency transformation of *Plasmodium falciparum* by the lepidopteran transposable element piggyBac. *Proc. Natl Acad. Sci. USA*, **102**, 16391–16396.
  39. Cary, L.C., Goebel, M., Corsaro, B.G., Wang, H.G., Rosen, E. and Fraser, M.J. (1989) Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology*, **172**, 156–169.
  40. Fraser, M.J., Cary, L., Boonvisudhi, K. and Wang, H.G. (1995) Assay for movement of Lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology*, **211**, 397–407.
  41. Fraser, M.J., Ciszczon, T., Elick, T. and Bauser, C. (1996) Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Mol. Biol.*, **5**, 141–151.
  42. Lopez, R., Silventoinen, V., Robinson, S., Kibria, A. and Gish, W. (2003) WU-Blast2 server at the European Bioinformatics Institute. *Nucleic Acids Res.*, **31**, 3795–3798.
  43. Mochizuki, T., Wu, G., Hayashi, T., Xenophontos, S.L., Veldhuisen, B., Saris, J.J., Reynolds, D.M., Cai, Y., Gabow, P.A. *et al.* (1996) PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science*, **272**, 1339–1342.
  44. Dai, B. and Mei, C.L. (2006) Research on autosomal dominant polycystic kidney disease in China. *Chin. Med. J. (Engl)*, **119**, 1915–1924.
  45. Beebe, D.K. (1996) Autosomal dominant polycystic kidney disease. *Am. Fam. Physician*, **53**, 925–931.
  46. Wu, G., Markowitz, G.S., Li, L., D’Agati, V.D., Factor, S.M., Geng, L., Tibara, S., Tuchman, J., Cai, Y. *et al.* (2000) Cardiac defects and renal failure in mice with targeted mutations in Pkd2. *Nat. Genet.*, **24**, 75–78.